PB85-198364

Data Base Development and Analysis for Water Distribution Systems

(U.S.) Environmental Protection Agency
Cincinnati, OH

May 85

U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before comple

1. REPORT NO.	2.	3. PB85-198364
EPA/600/D-85/098		

4. TITLE AND SUBTITLE	5. REPORT DATE
DATA BASE DEVELOPMENT AND ANALYSIS FOR WATER DISTRIBUTION SYSTEMS	May 1985
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
James A. Goodrich, Jeffrey Q. Adams, Robert M. Clark	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
Drinking Water Research Division USEPA, WERL 26 W. St. Clair Street Cincinnati, OH 45268	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Water Engineering Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Cincinnati, OH 45268	14. SPONSORING AGENCY CODE EPA/600/14

15. SUPPLEMENTARY NOTES

16. ABSTRACT

There have been a number of studies estimating the costs of repairing, replacing, or renovating our aging distribution networks. Although useful in general terms, characterizing a system by age and/or breaks per mile of mains is of limited use to a utility trying to manage increasing maintenance costs. The purpose of this paper is to (1) provide insight into developing a data base centered approach for reliability analyses and (2) suggest which analytical tools may be the most useful.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
		6
	20. SECURITY CLASS *(This page)*	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

EPA/600/D-85/098
May 1985

DATA BASE DEVELOPMENT AND ANALYSIS FOR WATER DISTRIBUTION SYSTEMS

by

James A. Goodrich,[1] Jeffrey Q. Adams[2]

and Robert M. Clark, M.ASCE[3]

ABSTRACT: There have been a number of studies estimating the costs of repairing, replacing, or renovating our aging distribution networks. Although useful in general terms, characterizing a system by age and/or breaks per mile of mains is of limited use to a utility trying to manage increasing maintenance costs. The purpose of this paper is to (1) provide insight into developing a data base centered approach for reliability analyses and (2) suggest which analytical tools may be the most useful.

INTRODUCTION

The U.S. Environmental Protection Agency has a variety of tasks underway that deal not only with the quality of drinking water within a distribution system but also with the costs and reliability of providing that water. The treatment and delivery functions of a water utility represent large economic investments, but the vast majority of expenditures are in the delivery system. Population projections made in designing water systems frequently become self-fulfilling. The water utility plays a significant role in community public health and can also become a determinant of the communities' growth path. Traditionally, most water utilities repair or replace mains as the situation dictates, not as they planned. Often, a hydraulic model is only used to test fire flow while one or two knowledgeable employees intimately know and understand the distribution system. Although utilities possess large amounts of data, it is rarely in a form easily accessible or useful in providing reliability analyses or developing decision-making criteria for repair or rehabilitation.

DATA BASE CENTERED APPROACH

Each utility is unique with its own set of records, distribution system problems and goals in developing a management tool. The use of a data base centered approach linking the distribution network spatially

---

[1]Research Environmental Scientist, Systems and Cost Evaluation Activity, Drinking Water Research Division, WERL, Cincinnati, OH 45268
[2]Environmental Engineer, Systems and Cost Evaluation Activity, Drinking Water Research Division, WERL, Cincinnati, OH 45268
[3]Director, Drinking Water Research Division, WERL, Cincinnati, OH 45268

with a "geo-code" facilitates multiple tasks. The key to drinking water supply delivery analysis is the performance of reliability assessments. The critical component of an inventory and water main condition assessment is the geo-spatial coding of data, thus providing a common reference system. The Louisville, Kentucky and Cincinnati, Ohio Water Works are undertaking projects which include (1) an overall description of the system (connectivity, physical characteristics, etc.), (2) structural integrity, (3) quality of service and level of usage, and (4) safety.[4] Each link, node, break and customer complaint can be uniquely indentified by its location either in State Plane Coordinates and/or the United States Bureau of Census Dime File code. In this manner, a variety of data such as soil type, land use, traffic load, operating pressure, elevation, etc., can be assessed. Reliability analyses using any model can also be integrated into the framework since the network is already specified. Thus, "reliability" analyses need no longer be simply simulations calculated by a hydraulic model, but can also incorporate socio-economic variables. The ability of a utility to anticipate, detect, repair, rehabilitate, reconstruct or replace water supply delivery components is enhanced by the geo-spatial coding of data which provides a common reference system.

DATA COLLECTION

Data collection, or more appropriately, data organization within most utilities is a painstaking process. The data are already present, usually in file cabinets, but of limited use in this form. The first step in developing a useful data base is to identify links and nodes for the system. A link should be defined as short segments, much shorter than what is typically used in the modeling of transmission mains. A link can be defined as a change in diameter, an intersection or even a run of pipe between valves. Useful information for each link should include both nodes, pressure, installation date, diameter, material, length, etc. Maintenance information should include specific street address of maintenance event, type of problem (main line leak, joint leak), depth, type of action taken, local environmental conditions, plus any other available information the utility desires.

DATA MANAGEMENT

Data entry, storage, retrieval and file manipulation using a spatial orientation on a microcomputer are best handled by a "relational" data base management system. In this type of data base, data are organized into a series of interrelated two dimensional tables. Such a system is easy to construct. Figure 1 is an example of how drinking water distribution system network data can be organized in a relational form. Information from any table can be used in conjunction with retrievals through relationships shown by the dotted line. Data entry is menu-driven without expensive on-line timesharing costs to a large mainframe computer. This saves money, time and effort while requiring little programming or computer expertise. This form of data organization also allows expansion of the individual data base tables so that new types of information can be included with the existing types of data.

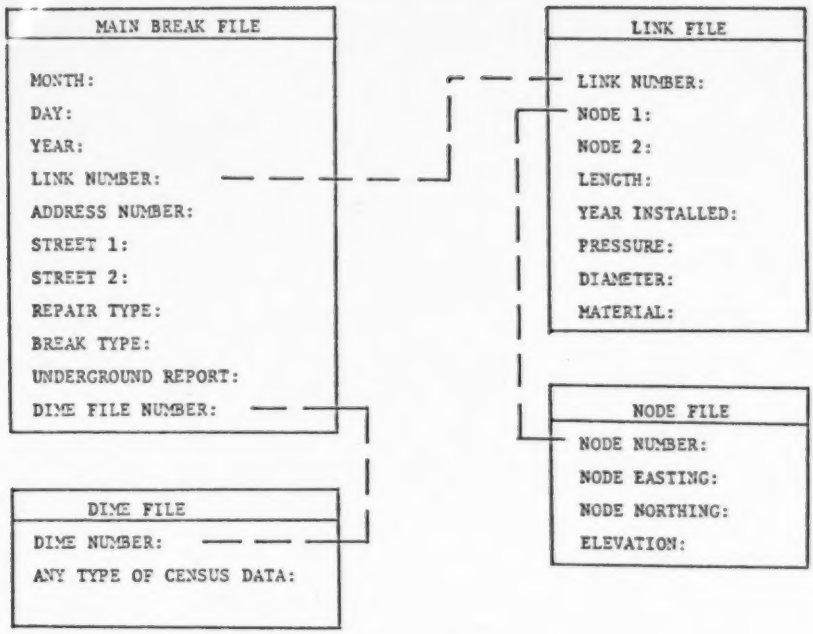

Figure 1. Relational Data Base Organization

DATA ANALYSIS

Microcomputer data base management systems vary in capabilities but generally are able to produce descriptive statistics, spreadsheets, graphics and english-like queries. A very useful aspect of a relational data base management system is its sorting capability. Sorting data by those mains with multiple maintenance events within a range of geographic reference points can identify "hot spots" or pipes of similar characteristics that could then be flagged as priority items for action. If a mainframe computer or sophisticated statistical software is available, a data file can be created to meet the model or program requirements through the sorting mechanism. Thus, optimization techniques, hydraulic analyses, network reliability analyses etc., can be performed without locking the data into a specific format only of use to one model.

Case Study Examples

In both the Cincinnati, Ohio and New Haven, Connecticut pipe break studies, analysis of the data showed that only a very few pipes examined caused most of the maintenance problems.[1,2] In fact, approximately two-thirds of the pipes have no maintenance events associated with them at all, even after very long periods of time. In addition to looking at

general historical records, regression analyses attempted to evaluate the various factors that influence the frequency of repair events. In order to quantify some of the relationships, general regression analysis was performed. For New Haven, significant variables related to breaks/mile of pipe included age, soil stability, operating pressure and whether the pipe was installed after 1960, suggesting the manufacturing technique, installation practice or class of pipe might be important to break rates. Before 1960 most distribution systems were thick wall cast iron pipe, after 1960 it was thinner wall. The thinner wall pipe may be breaking at an increasing rate as compared to the thick wall pipe installed before that data. Analysis for Cincinnati also included age, and post-1960 installation variables. Pressure and soil stability variables were replaced by soil corrosivity and pipe wall thickness.

Examination of the individual repair event history is critical. Even in high maintenance event areas, there is a wide variation among the individual pipes and their historical records. It appears that there are two time periods associated with maintenance events for individual pipes. Period one is from installation to the first repair. This event could often be the result of factors totally different from all other repairs throughout the pipe's history, but the time to this first repair often sets the tone for the remainder of the pipe's useful life. Generally, after each repair event, the time to the next event becomes increasingly shorter, thus, the sooner the first event, the sooner the next break and the higher the total break rate.

An interesting aspect discovered in both case studies is the variation in rates that occur over the delivery area. Break rates vary from 0 to 30 per mile in different census tracts. Different areas are subject to different causes of breaks thus limiting the usefulness of regression analysis over the entire service area. To combat this difficulty, regression analysis can be performed on sub-regions of the delivery system.

In addition to general deterministic regression approaches, probabilistic regression analyses were performed on the New Haven and Cincinnati pipe maintenance data.(3) The approach combined in-depth descriptive and bivariate statistical analysis with the regression analysis procedure developed by Cox in 1972. The Cox regression model is a general failure prediction model which was applied to the water distribution data concerning pipes that have had maintenance events and also those pipes that have not had maintenance events. The Cox regression model determines the probability of a pipe experiencing a maintenance event (break, leak) at any given point in time as a function of certain direct or surrogate factors that appear to be correlated to the pipe break mechanism. One of the purposes of performing this regression analysis was to define the variables affecting the probability of maintenance events and to determine the relationship between them. The major variables affecting the probability of pipe breaks included

internal pipe pressure, land development covering the pipe, age of pipe at the time of the second break, installation period, and the number of previous breaks on the pipe.

Cox's regression model derives the probability of experiencing a break in each specific pipe rather then the expected number of breaks in a given year as determined using the general regression approach. The Cox model is thus particularly well suited to the prediction of maintenance events for systems with relatively infrequent break events. Large mains tend to experience fewer breaks yet these breaks are of more extreme consequences. The prediction of these more consequential breaks is of great interest to utility management. The Cox regression model is designed to function on such pipes with relatively low probabilities of failure while several of the other deterministic regression models are less appropriate. The probability results from the Cox regression models can be directly used to prioritize maintenance actions. These results could also be combined with cost information associated with the different maintenance alternatives in order to rank pipes for maintenance strategies. This information is intended to be used as an aid to support the organization and preparation of water distribution maintenance plans and budgets.

APPENDIX A. REFERENCES

1. Clark, R. M., Stafford, C. L., and Goodrich, J. A., "Water Distribution Systems: A Spatial and Cost Evaluation", Journal of the Water Resources Planning and Management Division, ASCE, Vol. 108, No. WR3, Proc. Paper 17425, October, 1982, pp. 243-256.

2. Clark, R. M., and Goodrich, J. A., "Repair, Renovation and Replacement Strategies in Drinking Water Systems", presented at the May 28-31, 1984, ASCE Water Resources Planning and Management Annual Conference, Urban Water '84, held in Baltimore, MD.

3. Marks, D., Jeffrey, L., Andreo, S., and Park, C., "Predicting Urban Water Distribution Maintenance Strategies: A Case Study of New Haven, Connecticut", EPA Draft Report, 1985.

4. O'Day, D. K., and Neuman, L. A., "Assessing Infrastructure Needs: The State of the Art", Perspectives on Urban Infrastructure, Royce Hanson, ed., National Academy Press, Washington, D.C., 1984, pp. 72-75.

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.